Relating soybean (*Glycine max)* Ecotype and Environment to Seed Biomass Composition

Undergraduate Research Thesis

Presented in partial fulfillment of the requirements for graduation
with research distinction in Biochemistry in the undergraduate colleges of
The Ohio State University

by

Tamara McClure

The Ohio State University
April 2015

Project Advisor: Dr. Ana Alonso, Department of Molecular Genetics

Background

The Soybean plant (*Glycine max* (L.) Merrill) is a species of legume that produces pods that contain beans, which are referred to as seeds when mature. Current soybean cultivars are much different genetically from wild-type soybean plants (*Glycine soja*) due to selective pressures from thousands of years of agricultural production [17,18]. The fresh beans from green pods can be eaten as a vegetable, and dry mature seeds are used in animal feeds and human foods that contain soybean oil or meal, and in industry in solvents, lubricants, inks, plastics, waxes and others (**Figure 1)** [9]. The soybean plant's market value stems largely from the significant oil and protein content of its seeds which account for approximately 40% [20] and 20% [5] of total dry weight, respectively, giving the seeds appreciable versatility. Because of these attributes, soybean accounts for a significant amount of the world's vegetable oil, animal fodder and food for human consumption.

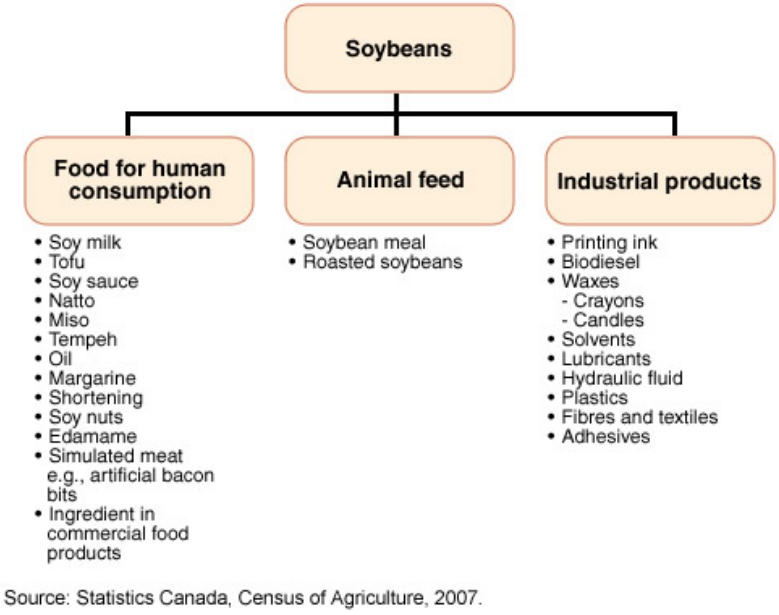

Figure 1: General uses of soybeans [9].

Soybean is an integral part of the diet of many Asian cultures with uses in items such as soymilk, soy sauce, soy paste, edamame, tempeh, miso, tofu and natto [9]. Among western cultures soybean is used mostly for its meal and oil. In 2013, 49.8% of soybeans produced in the United States were crushed for oil and meal, 2.8% was used for seed, feed and other purposes, and the remaining 47.4% was exported. Soybean exports have increased by about 33.5% from 2000 to 2013 from 996 to 1647 million bushels, for a total price of $21.4 billion in 2013 [35]. The United States is a top soybean producer along with Brazil followed by Argentina, China and India. According to the USDA, soybeans comprise about 90% of U.S. oilseed production and make up the world's largest protein source in animal feed and second largest source of vegetable oil [34], emphasizing not only the United States' role as a major producer, but also the global demand for soybean. US soybean production is the highest in Midwestern states including Illinois, Iowa, Minnesota, Indiana, Nebraska and Ohio [8]. Research regarding soybean is relevant in the context of these major areas of production.

In order to maximize utility and market value of soybeans, plant breeding programs and genetic modification efforts aim to improve nutritional value. Scientific research efforts investigate naturally produced compounds, genetics, and metabolic pathways in crops for the benefit of these long-term goals. In the case of soybean, the content and composition of biomass components such as proteins and oils are targeted in research due to their marketability. Increased oil content, for example, is pursued in soybean to meet the demands of the vegetable oil market. Soybean oil is comprised of fatty acids such as oleic acid (18:1, approx. 18% of crude oil), and essential linoleic (18:2 omega-6, 55% of crude) and linolenic (18:3, omega-3, 13% of crude) fatty acids, which are necessary for health and must be obtained by consumption [5]. Though essential and desirable, linolenic and linoleic acids are responsible for oxidative

instability of soybean oil, which has been historically addressed through partial hydrogenation [5, 20]. Once hydrogenated and refined, soybean oil is a good source of vegetable oil, however this is at the cost of health, due to the resulting trans-fats that are shown to increase the risk of coronary heart disease [20]. Naturally high oleic acid content in soybean is valuable due to this monounsaturated omega-9 fatty acid's stability at high temperatures, which reduces the production of trans-fats during vegetable oil production and use [20]. For the same reason, low linolenic soybean is also pursued to decrease the amount of oxidative instability.

High protein soybean cultivars have also been pursued to meet market demands. Regarding protein composition, current agricultural efforts aim to increase the content of essential, limiting sulfur-containing amino acids critical to animal health, specifically, cysteine and methionine [16, 36]. Additionally, though increased total oil content and increased total protein both have been pursued independently in some soybean plant lines, current efforts aim to simultaneously increase the contents of both oil and protein. Several studies have shown that these components are inversely correlated [5, 18] therefore maximizing both will require alternative strategies. Due to soybean's versatility, there are various ways its seed composition can be improved. It would be greatly beneficial to the various industries that manufacture soybean products to develop plants with adequate nutrient compositions to reduce post-harvest processing and increase marketability.

Both breeding programs and genetic modification efforts are used for the purpose of creating specialized soybean lines. Successful efforts include the creation of soybean lines with increased target nutrients, and lines with resistances to herbicides, pesticides, and pathogens [19, 28]. Breeding has been a widely employed strategy for improving soybean for over two thousand years, and genetically modified (GM) soybeans have been commercially grown in the U.S. since

1996 [13]. Breeding and GM efforts often rely on biochemical surveys of soybean seeds to identify seeds with genotypes that produce desirable phenotypes to use for their purposes.

Genotype and environmental factors directly affect biomass content in plants. In order to develop product-specific soybeans, these major factors must be properly investigated and understood. In addition to genetics, environmental factors such as water and soil nutrient availability, soil pH, light, plant density and temperature directly affect the production of plant biomass components as well [21, 26, 29]. Determining correlations between biomass composition and both genotype and environmental factors in soybean will create avenues for genetic engineering efforts and allow for the adjustment of agricultural practices in order to optimize the production of product-specific soybean lines. To this end, this study was pursued in order to initiate the development of a database of the biomass composition in various soybean cultivars grown at different locations to aid agricultural efforts, genetic engineering efforts, and breeding programs.

We hypothesize that there are differences in biomass composition that can be directly attributed to differences in either genotype or environment, or both. To test our hypotheses, ten different Ohio-adapted soybean cultivars (**Table 1**) were grown at four locations in Ohio that covered three different soil types (**Figure 2**).

Table 1. List of soybean cultivars that were chosen in this study.
Highlighted: Four cultivars chosen for further statistical analysis.

Cultivar	Classification	Status
Dennison	Commodity	Public Cultivar
H09-4	High Protein	Breeding Line
M09-W043	High Protein	Breeding Line
M09-W053	Commodity	Breeding Line
M09-W150	Commodity/ Mod. Fatty acid	Breeding Line
M09-W153	Food Grade	Breeding Line
Ohio FG1	Food Grade	Public Cultivar
Ohio FG5	Food Grade	Public Cultivar
Summit	Commodity	Public Cultivar
Wyandot	Commodity/ Food Grade	Public Cultivar

The cultivars were grown in the Ohio locations Plain City (PC), South Charleston (WE), Hoytville (NW), and Wooster (WO) (**Figure 2**). Three of the growth locations, WE, NW, WO, are Ohio Agricultural Research and Development Center (OARDC) farms and employed the similar agricultural practices whereas PC is non-OARDC.

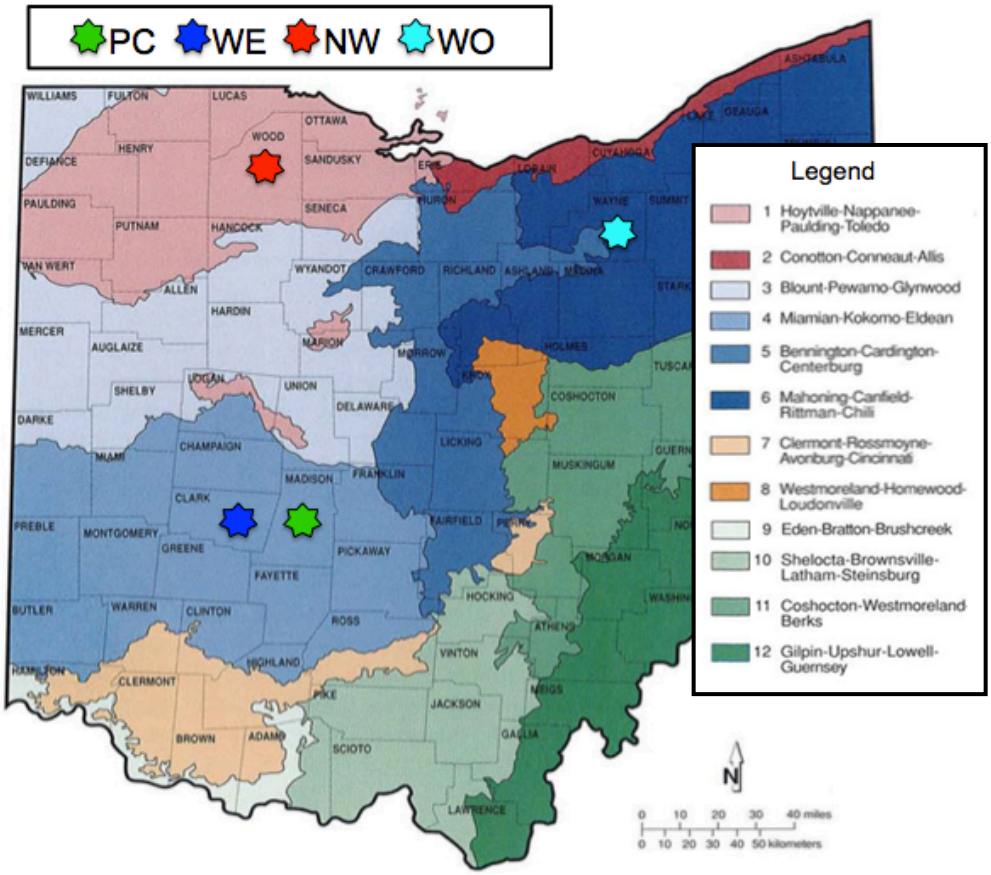

Figure 2: Soil regions in Ohio and growth locations under study

Retrieved from: www.soilandwater.ohiodnr.gov

Materials and Methods

Materials

Mature seed samples were obtained from Dr. Leah McHale (Dept. of Horticulture and Crop Sciences, OSU) who has a collection from various cultivars and breeding lines grown in different locations in Ohio. For this project, ten cultivars were obtained from the 2013 growing season: Dennison, H09-4, M09-W043, M09-W053, M09-W150, M09-W153, Ohio FG1, Ohio FG5, Summit and Wyandot (**Table 1**). These cultivars were tested following growth at four locations, Plain city (PC), South Charleston (WE), Hoytville (NW), and Wooster (WO) (**Figure 2**).

Biomass extraction

Mature soybean seeds were dried in a lyophilizer and weighed to obtain the initial dry weight. Fatty acids and proteins were extracted from 50 mg of biological material as previously described [7] with the following modifications: The fatty acid extractions were carried out with 950 µL hexane:isopropanol (2:1) and 50 µL of triheptadecanoin (10 mg/mL) as an internal standard. Proteins were extracted using 1 mL of extraction buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, and 1% SDS). The process was repeated twice for a total final volume of 3 mL of protein-containing supernatant to be used for quantification.

Biomass quantification

Total oil content and oil composition of each seed sample was determined by Gas Chromatography-Mass Spectrometry (GC-MS) analysis of fatty acid methyl esters. The oil fraction containing triheptadecanoin (17:0) internal standard was transmethylated into fatty acid

methyl esters as previously described [7] with some modifications. Fatty acids were methylated instead using 300 μL of toluene and 1 mL of 3N methanolic/HCl. Transmethylation was carried out for 120 minutes at 80^{o}C, and the reaction was quenched using 500 μL 5% (w/v) sodium bisulfate. Oil content was quantified using GC-MS analysis as previously described [33] using a run time of 6 minutes. Protein concentration of each sample was measured using the Biorad DC Protein Assay kit, which employs a modified Lowry assay for colorimetric protein quantification.

Proteinogenic Amino Acids

To determine the composition of amino acids, the protein extracts were hydrolyzed using 6N HCl for 24 hours at 120°C. Free proteinogenic amino acids obtained from hydrolysis were purified through a cation exchange resin (Dowex 50X8) and then analyzed by Liquid Chromatography tandem Mass Spectrometry (LC-MS/MS) using methods as previously described [7]. The quantification of each amino acid was possible by running known concentrations of external standards. The quantitation method included sixteen amino acids that are naturally found in soybean seeds: Alanine, Arginine, Aspartate, Glutamate, Glycine, Histidine, Methionine, Leucine, Isoleucine, Lysine, Phenylalanine, Proline, Serine, Threonine, Tyrosine and Valine.

Statistical Analyses

For each biomass component under study, the mean and standard deviation of the four biological replicates were calculated for each location and each cultivar. To determine statistical significance, a student's t-test was to compare each parameter under study with respect to

location. P-values of < 0.05 were considered significant. Supervised multivariate partial least-squares discriminate analysis (PLS-DA) was performed using Metaboanalyst v2.5 (Xia *et al.* 2009, 2012), a free web-based statistical software (www.metaboanalyst.ca). PLS-DA showed to what degree cultivars and locations separated with respect to the components identified. These components were determined through dimensional reduction and redefining the axes to account for all variables under study. Data were pre-processed for PLS-DA using median normalization to create a normal distribution and generalized logarithm transformation to stabilize variance. PLS-DA allowed for the visualization of differences amongst both locations and cultivars, independently, as well as variance. The separation of the groups visualized in the graphs generated by PLS-DA analysis reflected these differences and the spread of the data points amongst each group, location and cultivar, reflected the variance. Variable Importance in Projection (VIP) analysis was also performed, where VIP is a "weighted sum of squares of the PLS taking into account the amount of explained Y-variation in each dimension" [37, 38]. VIP scores determined the metabolites that counted for the most separation for both cultivars and locations.

Results

Study of 10 Ohio-adapted Soybean Cultivars

This entire study incorporated a total of ten cultivars. In order to perform the calculations necessary to quantify and characterize biomass components of the samples under study, the dry weight for each biological replicate seed was measured and the average dry weight was calculated. Oil, proteins and the relative amount of measurable amino acids were determined in order to characterize biomass composition. **Figure 3** below shows the dry weights (a), total oil

contents (b), protein contents (c), and total amount of essential amino acids (d) for all ten cultivars grown at all four locations. The total amount of essential amino acids was obtained through a summation of the relative amounts of the essential amino acids measurable using the methods described above in conjunction with LC-MS/MS analysis: Histidine, Isoleucine, Leucine, Lysine, Methionine, Phenylalanine, Threonine, and Valine. For all samples, the relative amounts of all sixteen amino acids were calculated.

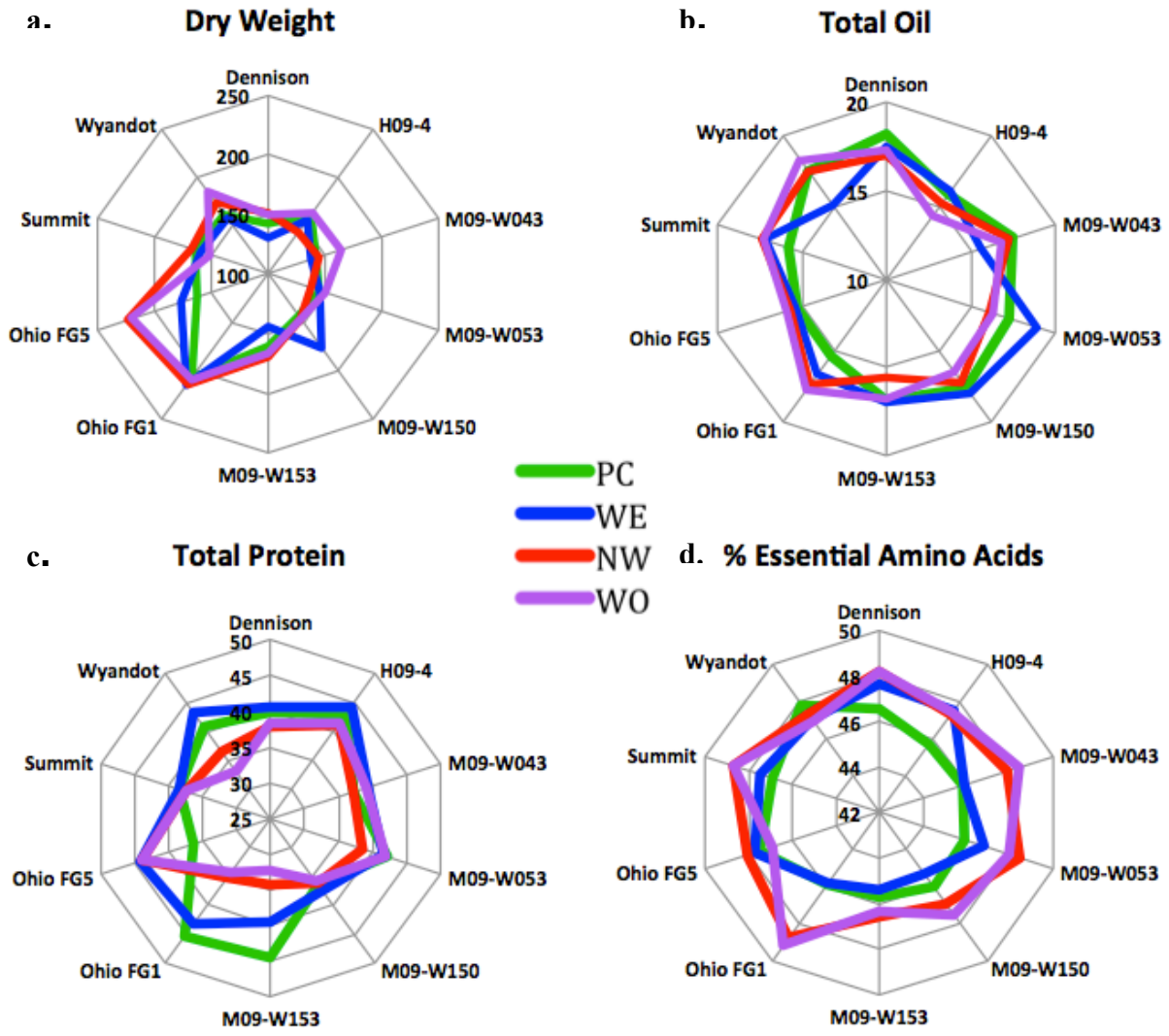

Figure 3. Biomass characterization of 10 soybean cultivars grown in 4 different locations: (a) Average total dry weight (mg/seed; average of 20 seeds) (b) Average total oil (% dry weight, average of 4 biological replicates) (c) Average total protein (% dry weight, average of 4 biological replicates) and (d) Average total essential amino acids (% total amino acids, average of 4 biological replicates).

Out of all of the ten cultivars, three of the four food grade soybean cultivars under study, Wyandot, Ohio FG1 and Ohio FG5, generally had seeds with the highest dry weights. However the dry weights for Ohio FG5 were significantly higher at the NW and WO locations compared to both PC and WE locations (**Figure 3, Table A-2**). Additionally, the average dry weight for M09-W150 seeds at the WE location was significantly higher than that of the PC, WE, and WO locations ($p<0.05$, **Figure 3, Table A-2**). Protein content was highest for soybean seeds from the Wyandot, Ohio FG5, Ohio FG1 and M09-W153 cultivars, and generally dependent on the location. The M09-W053 cultivar had the highest oil content and there was significant variation in total oil content with respect to location for both M09-W053 and Wyandot cultivars (**Table A-6**). There was considerable variation regarding the total amount of essential amino acids across all of the cultivars. Soybean seeds from the WO and NW locations generally had higher amounts of essential amino acids, significantly so for the Ohio FG1, M09-W150, M09-W053 and M09-W043 cultivars. Seeds grown in the PC location generally had the lowest amount of essential amino acids.

Statistical Study of 4 Ohio-adapted Soybean Cultivars

In order to better visualize differences in biomass composition, four of the original ten cultivars were chosen based on preliminary PLS-DA analyses in order to make general observations and conclusions (**Figure 4**). Among all ten

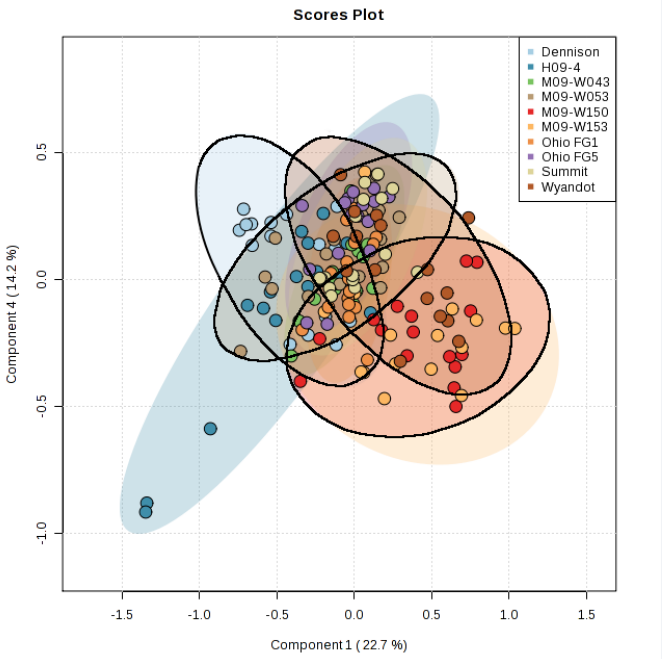

Figure 4. PLS-DA for all ten cultivars and the 95% confidence regions for the four chosen cultivars.

cultivars there is significant overlap and it is difficult to see where the differences lie. Dennison discernably separated from both Wyandot and M09-W150 cultivars and M09-W053 and M09-W150 separated from each other as well. Additionally with respect to location, M09-W053 and M09-W150 were variable in both fatty acid composition and amino acid composition, Wyandot was variable in total protein content and fatty acid composition and Dennison was fairly stable across all locations. For these reasons these four cultivars were chosen for closer investigation in this study.

Dry Weights Among Four Cultivars

The dry weights for the four cultivars chosen were more closely examined (**Figure 5)**. Out of the four cultivars tested, Wyandot had the highest dry weight compared to the other cultivars at the PC, NW and WO locations. Growing location had less of an impact on the M09-W053 seeds than the other cultivars. No growth location consistently yielded the largest or smallest seeds.

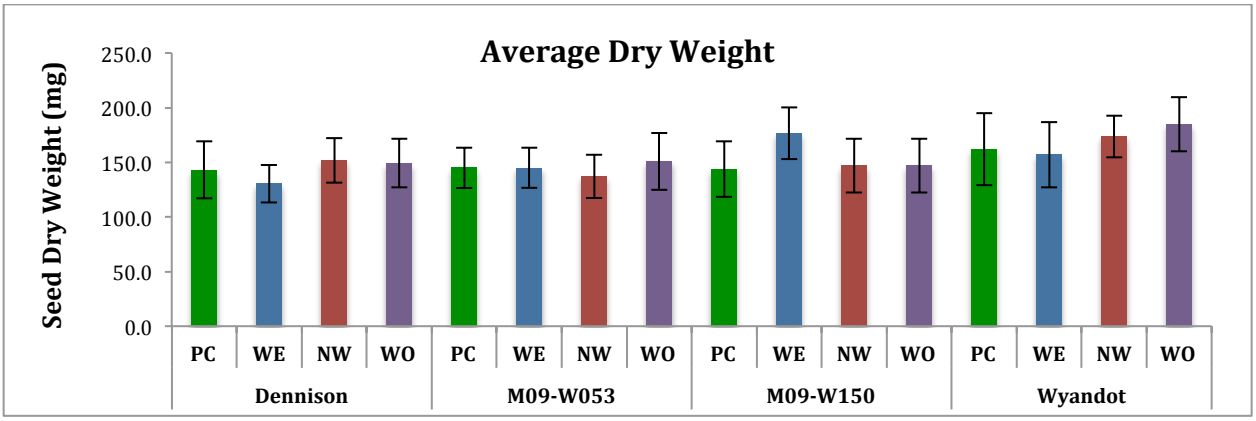

Figure 5. Dry weights of seeds from the four cultivars grown at the four locations. Average of 4 biological replicates ± SD.

Biomass Composition

Total oil and protein contents were determined for the four biological replicates obtained from each cultivar grown at each location. Results were calculated as percentages of total dry weight

(w/w) and the averages of the four biological replicates are reported along with standard deviation for each group (**Figure 6**). Total oil ranged from 15.2 to 18.0% and total protein ranged from 33.3 to 42.3%. There was significant variation in protein and oil content for both M09-W053 and Wyandot cultivars with respect to location ($p<0.05$, **Table A-6**).

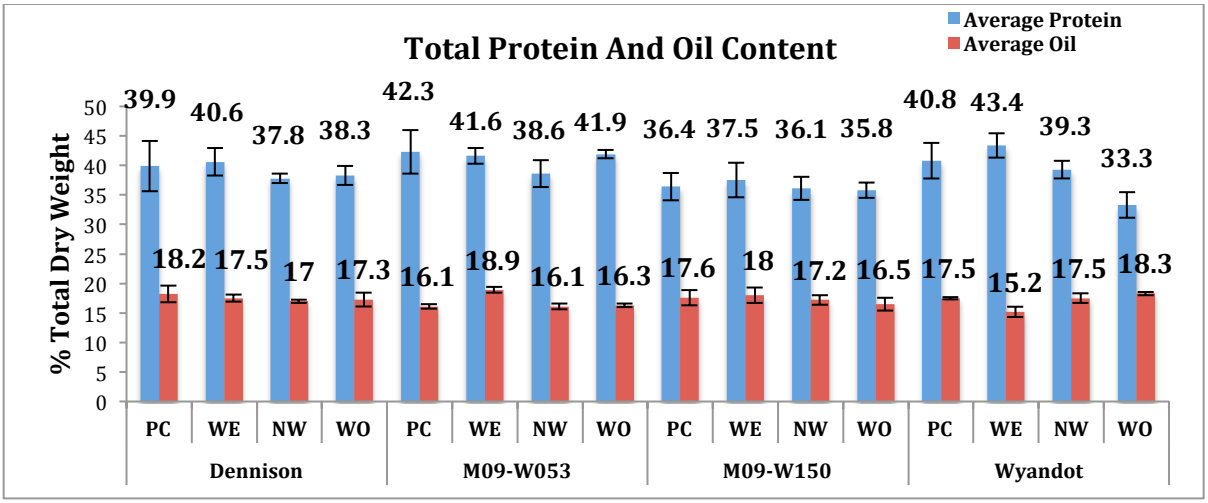

Figure 6. Total protein and oil content for each of the four cultivars grown at four locations

There was larger variation in total protein among all samples as compared to total oil. ANOVA shows there was no significant difference among total oil content across cultivars but confirms a significant difference in total protein content ($p= 0.028$). **Figure 7** revealed that there was only a weak negative correlation between total oil and total protein content ($r^2=0.133$).

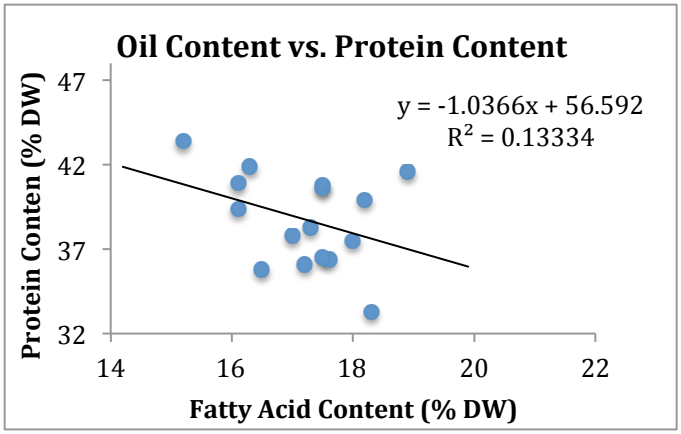

Figure 7. Correlation between protein and oil contents. Average total protein content (% of total dry weight) was represented as a function of average total oil content (% of total dry weight) for the four cultivars grown at four locations.

Five fatty acids derived from triacylglycerols were found in all samples: palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1), linoleic acid (C18:2) and linolenic acid (C18:3). The presences of these have been previously well documented in soybean [5]. There was variation in both fatty acid and amino acid compositions with respect to both cultivar and location (**Table A-5, A-7**). As previously acknowledged, sixteen amino acids were present in all samples, all of which have been well documented in soybean [31]. Cysteine, tryptophan, asparagine and glutamine are also present in soybean protein but could not be measured due to the acid hydrolysis. Cysteine and tryptophan were degraded during hydrolysis and asparagine and glutamine were converted to aspartate and glutamate, respectively.

Statistical analysis

Partial-Least Squares Discriminate Analysis (PLS-DA) allowed for the visualization of the extent to which cultivars and locations differed from each other. Each data point represents one biological replicate –one seed of a specific cultivar grown at one of the four locations. There was little separation between locations (**Figure 8a**) but there was discernable separation between cultivars (**Figure 9a**). Specifically, Wyandot and M09-150 cultivars both separated noticably from Dennison and M09-W053. Dennison and M09-W053 cultivars overlaped significantly. There was also more variance among locations than culivars, as visualized through the spread of the data points. Regarding locations, NW and WO had the most variance and PC had the least. For cultivars, Dennison had the least amount of variance.

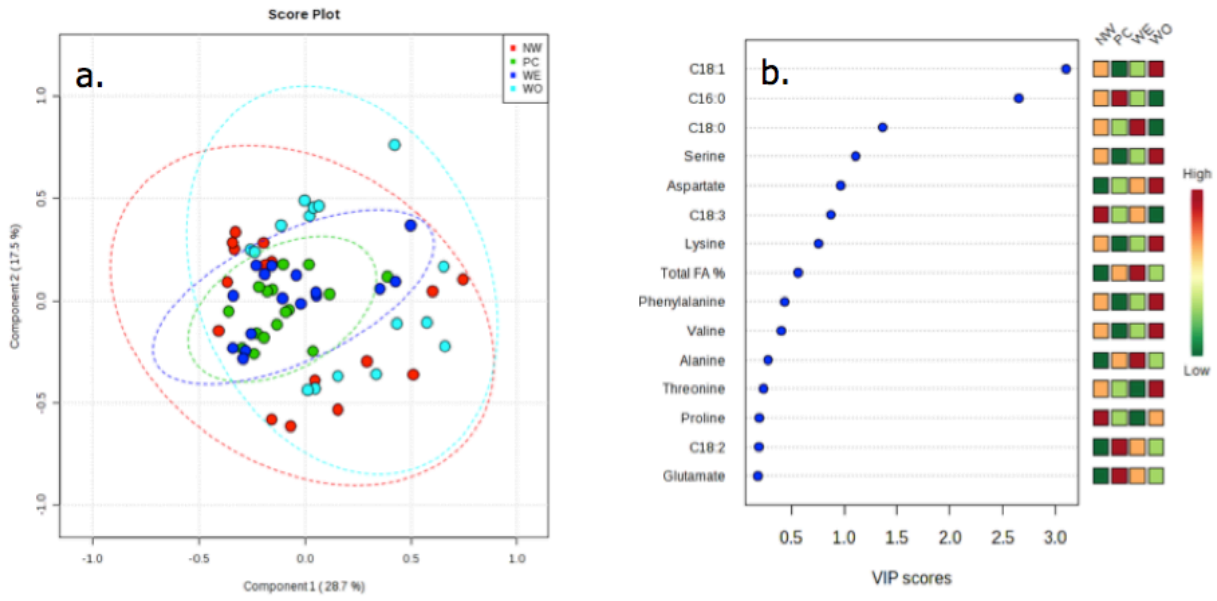

Figure 8. (a) PLS-DA score plot of component 1 (28.7% of total variance) and component 2 (17.5% of total variance) of the metabolite profile differentiating the four locations from which samples from four cultivars were harvested. (b) List of the important compounds displayed according to their score and weight for each soybean growth location.

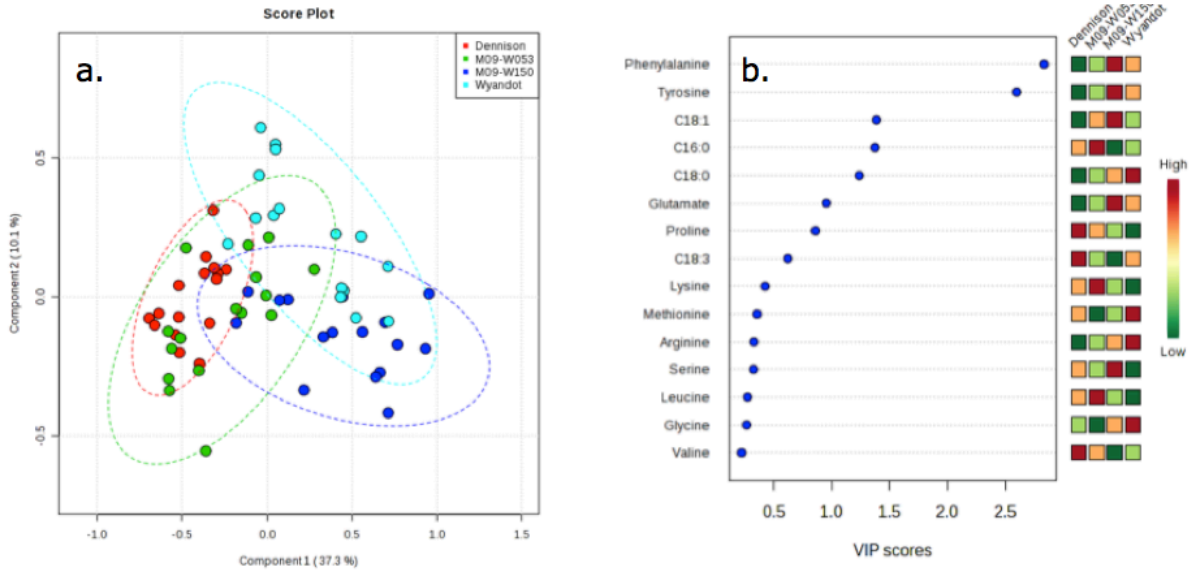

Figure 9. (a) PLS-DA score plot of component 1 (37.3% of total variance) and component 2 (10.1% of total variance) of the metabolite profile differentiating the four soybean cultivars grown at harvested from four growth locations. (b) List of the important compounds displayed according to their score and weight for each soybean growth cultivar.

VIP scores showed the primary factors responsible for differences between both locations and cultivars, independently of the other. Fatty acids accounted for more separation and variance among locations while amino acids accounted for more separation and variance among cultivars. Oleic acid (C18:1) and palmitic acid (C16:0) were the metabolites most responsible for differences among locations (**Figure 8b**) and phenylalanine and tyrosine were the metabolites most responsible among cultivars (**Figure 9b**). The heat maps associated with the VIP analysis showed which location or cultivar, in each case, had relatively more or less of that compound compared to the other locations or cultivars. These trends were validated with the histograms for each of the important compounds identified by VIP (**Figures 10 and 11**). Both the heat map legend provided by VIP and the relevant histogram show that palmitic acid was generally higher at the PC location and lower at the WO location (**Figure 10**). Conversely, oleic acid was generally higher at the WO location and lower at PC, suggesting that the two fatty acids were inversely correlated. Regarding phenylalanine and tyrosine content across cultivars, both were highest for M09-W153 and lowest for Dennison. The same observation could be made for the locations, suggesting a positive correlation between the two amino acids.

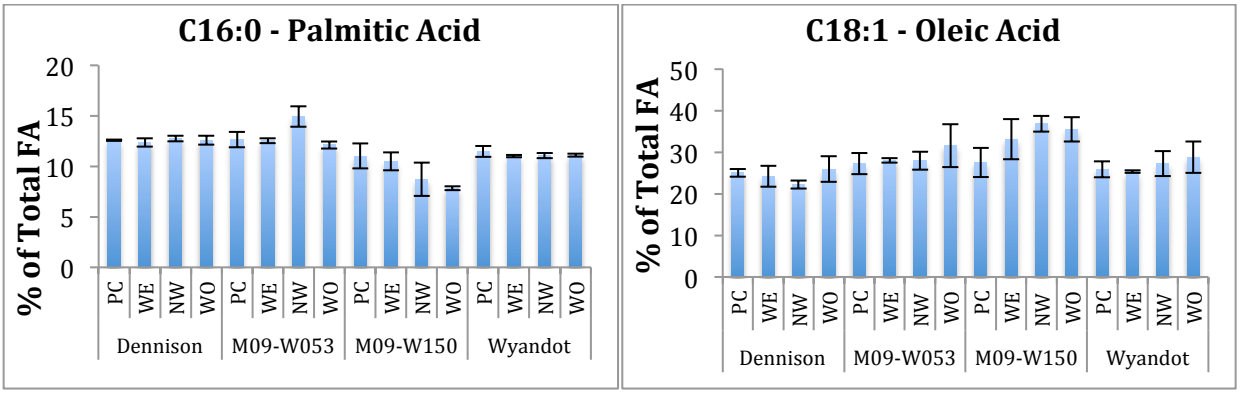

Figure 10. Contents of palmitic (C16:0) and oleic (C18:1) acid, important compounds identified by VIP for locations. Values are percentages of total FA (w/w) and average +/- standard deviations of 4 biological replicates.

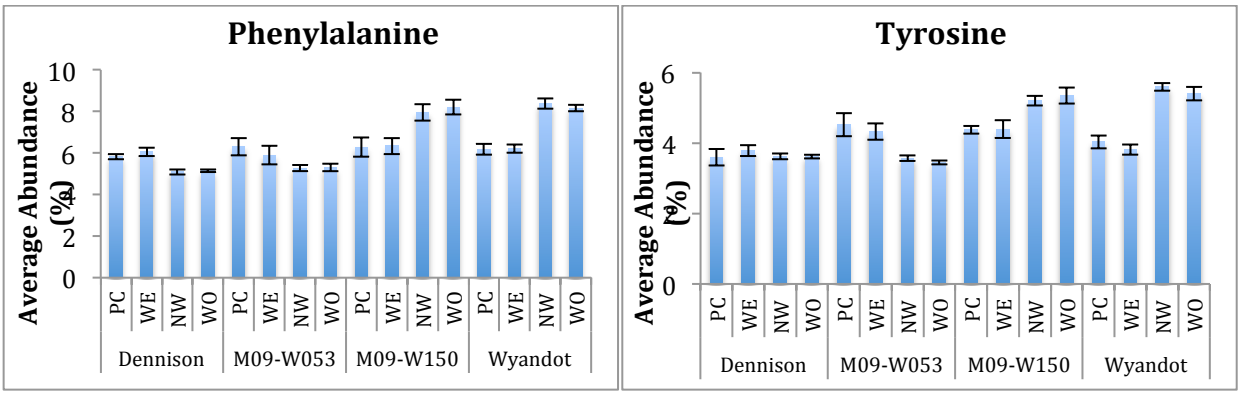

Figure 11. Abundances of phenylalanine and tyrosine, important compounds identified by VIP for cultivars. Values are percentages of total amino acid content and average +/- standard deviations of 4 biological replicates.

Discussion

The purpose of this study was to characterize the biomass composition of soybean seeds from several genetically different soybean cultivars grown at four locations. To this end, the techniques and analyses used in this study allowed for an efficient processing of the samples and analysis of data. GC-MS analysis quantitatively characterized free fatty acids derived from triacylglycerols, the main lipids stored in soybean seeds [20]. Inclusion of C17:0 internal standard ensured accuracy by accounting for any sample lost during sample preparation provided the mixture was kept homogenous. LC-MS/MS analysis qualitatively characterized amino acid composition by determining the relative amount of each amino acid measured. An external standard was prepared using known concentrations of each of the amino acids present in soybean and run as a calibration tool. Instrument responses from the target compounds in the samples were quantified according to the known external standards. Cysteine, tryptophan, asparagine and glutamine are present in soybean protein but were unmeasurable in this study due to effects of the protein hydrolysis used to free amino acids prior to LC-MS/MS analysis. Cysteine and tryptophan were destroyed and asparagine and glutamine were deamidated into their respective acids, inflating aspartate and glutamate contents accordingly [31].

Regarding the statistical analyses that were employed by this work, PLS-DA is one of two multivariate analyses that is used increasingly often in metabolomics studies. "A PLS model will try to find the multidimensional direction in the x space that explains the maximum multidimensional variance direction in the y space" (Liu, R. 2008). This supervised analysis takes into account all of the variables under investigation as well as the classification provided, in this case cultivar or location. PLS-DA analysis showed that there was more separation between the cultivars compared to the locations under study. It can be inferred that differences in biomass composition are likely more influenced by genotype rather than environmental factors such as temperature, soil composition and agricultural practices, all of which are a function of growth location. There was also more variance among locations than cultivars. This supports the idea that environmental factors were not the major contributors to differences in biomass composition. There was, however, also a notable amount of variance among cultivars so it is therefore difficult to draw definite and specific conclusions from the PLS-DA data alone.

VIP analysis in this study showed that for locations and cultivars, different metabolites influenced changes in biomass in each case. Fatty acids and amino acids (specifically phenylalanine and tyrosine) accounted for more variance among locations and cultivars, respectively. The amount of oleic acid accounted for differences in both cases. This is notable due to the increased value of soybean seeds that are naturally high in oleic acid, which have oil of less oxidative instability. Regarding phenylalanine and tyrosine, though these two amino acids are not currently targets of breeding programs and genetic modification, phenylalanine is an essential aromatic amino acid that must be obtained through the diet, and tyrosine is conditionally essential [32]. Metabolites identified by our VIP analyses may be targets of future soybean breeding and genetic modification efforts in order to develop crops to meet specific

health needs. In the case of the phenylalanine, for example, there is a disorder called phenylketonuria where those affected cannot process phenylalanine [4]. Similarly, there are other documented amino acid disorders that also lead to serious health problems [15]. As other nutritional needs and health concerns become better understood, it will be beneficial to develop food crops with adjusted amounts of specific fatty acids or amino acids to meet the needs of individuals as well as the population.

According to this study, genetic factors, rather than environmental factors, play more of a role in differences in biomass composition. This observation emphasizes the practicality and effectiveness of breeding and genetic modification in optimizing the development of product-specific soybean lines. Specialized soybean lines have been most commonly developed through breeding efforts, agrobacterium-mediated transformation and through particle-bombardment [28]. Breeding efforts screen for naturally resistant soybean plants and select for target genes through backcrossing, single pod descent, pedigree breeding and bulk population breeding methods [28]. In soybean, such methods have been applied successfully to develop plants resistant to pests and pathogens such as phytophthora stem and root rot (*P. sojae*), Soybean Aphid and Soybean Cyst Nematode [19, 28]. Soybean oil composition has been successfully altered and total oil and protein contents have been increased through breeding efforts and by exploiting the natural variation found in the soybean germplasm [5, 28]. Efforts to alter soybean oil composition primarily focus on reducing oxidative instability by increasing oleic acid content and decreasing linolenic acid. Mid to high oleic acid soybeans (30-70% of total oil) have been developed through breeding efforts though this has been at the expense of yield [3, 24, 25]. Additionally, low linolenic acid soybean oil genotypes have been developed through mutational breeding, using chemical mutagenesis to create plants that produce seeds

with desired traits [10, 25]. Finally, with regards to improving seed composition, our study did not show the strong inverse correlation between oil and protein content suggested by other studies [5, 18]. It may therefore be possible to increase total oil content without affecting protein content and vice versa.

Soybean has also been successfully transformed using both agrobacterium-mediated and particle bombardment-mediated transformations. Agrobacterium-mediated transformation has been efficient in transferring multi-gene or large inserts. Resistance to the herbicide glyphosate in engineered soybean lines is one prominent example of this technique [17]. Transformation of soybean by particle bombardment has also been successfully employed too, and current research efforts aim to continue improving this method [17]. The practical applications and wide-use of breeding and genetic manipulation efforts for crop improvement emphasize the importance of studies such as this one, which will find biomarkers to facilitate the success of these efforts.

Future Work

There is a growing demand for product-specific soybean lines in order to efficiently produce soybean seeds with biomass compositions that reflect the market needs [17, 28]. It will be useful to have a database of soybean biomass data corresponding to a variety of different cultivars and locations to meet the needs of future potential studies that may target biomass components, including potentially important metabolites with implications that may not yet be fully known or understood. Future work would thus extend this study to incorporate more cultivars and more locations in order to create a more comprehensive database for the benefit of soybean breeding and genetic modification efforts. Once a comprehensive database of biomass composition as a function of both genotype and environment has been established, the data can be used to select

the cultivars to be grown at any given location. It will also be valuable to use this database to elucidate the genetic and biochemical basis for the trends observed to aid in the identification of targets for breeding and genetic modification efforts.

As far as other future directions, there are other variables such as annual temperatures that were not considered in this study and may also account for differences or confound underlying trends. While temperature is a function of growth location, any given year may have abnormal weather patterns. It thus will be beneficial extend this study over several years to account for temperature and weather patterns from year to year. Additionally statistical analyses that characterize genotype-by-environment interactions should be employed to investigate more how genotype and environment affect each other in conjunction.

Other factors that affect biomass composition should be investigated such as pathogens and climate change. Regarding pathogens, benzothiadazole is a chemical inducer of systemic acquired resistance that has been used to activate resistance to pathogen infection in Arabidopsis, tobacco, wheat, sunflower, rice corn and cotton in order to study and control crop diseases [12]. Benzothiadazole treatment is currently being studied as a means of simulating *P. sojae* infection in soybean. It will be valuable to determine impacts of this treatment on soybean metabolism and biomass composition to understand the effects of *P. sojae* without requiring the presence of the pathogen. The effects of climate change due to global warming are also a major concern of scientists, the agricultural industry, and the government, due to the potential threat on vital crops such as soybean. Other future directions of this study thus could incorporate studying the effects of drought and elevated atmospheric CO_2 to estimate the impact of climate change in current Ohio-adapted soybean cultivars.

Acknowledgements

I would like to thank Dr. Ana Alonso for giving me the opportunity to conduct research in her lab and for being my advisor, and I thank her for supporting and guiding me throughout my experimental process and the writing of my thesis. I thank Jean-Christophe Cocuron for giving me guidance and assistance regarding the experimental procedures used as well as the training on the instrumentation needed for this study, and thank the Targeted Metabolomics Facility at OSU for allowing me access to this instrumentation. I would also like to thank the other Alonso lab members, Enkhtuul Tsogtbaatar, Zack Ross, Brooke Anderson, Erin Ponting and Davesh Chauhan for their input and assistance, and visiting scholar Veronika Osmark for her technical help and contribution with this project. I acknowledge the Ohio Soybean Council for funding this research and the Ohio State Arts and Sciences Honors Committee for awarding me with an undergraduate research scholarship to aid in my studies as I carried out my research. Finally, I thank Dr. Jennifer Ottesen for her time and help as a member of my thesis committee and Dr. Leah McHale for providing the soybean seeds used for this project.

Appendix

A-1. Average Dry Weights for Ten Cultivars

		Average Dry Weight (mg)	Std. Dev.
Dennison	PC	143.0	26.1
	WE	130.4	17.1
	NW	151.9	20.5
	WO	149.4	22.3
H09-4	PC	161.3	23.9
	WE	155.7	31.0
	NW	142.3	19.5
	WO	163.2	16.9
M09-W043	PC	141.9	22.6
	WE	138.3	22.8
	NW	144.9	17.1
	WO	165.1	26.1
M09-W053	PC	145.1	18.6
	WE	145.0	18.3
	NW	137.2	19.8
	WO	151.0	26.1
M09-W150	PC	143.8	25.4
	WE	176.7	23.7
	NW	146.9	24.5
	WO	147.0	24.7
M09-W153	PC	159.7	27.1
	WE	144.3	27.0
	NW	168.7	22.3
	WO	166.6	14.5
Ohio FG1	PC	207.4	31.5
	WE	214.7	22.9
	NW	214.1	34.6
	WO	209.3	29.9
Ohio FG5	PC	161.8	45.5
	WE	175.8	40.3
	NW	223.8	28.9
	WO	219.8	26.0
Summit	PC	163.2	25.0
	WE	158.6	21.2
	NW	167.4	19.8
	WO	150.8	22.7
Wyandot	PC	162.0	32.9
	WE	157.0	29.9
	NW	173.6	19.0
	WO	184.9	24.8

A-2. p-values for Dry Weight Differences in Two Notable Cultivars

M09-W150	p-value
PC/WE	0.0001
PC/NW	0.696
PC/WO	0.693
WE/NW	0.0004
WE/WO	0.0004
NW/WO	0.994
Ohio FG5	**p-value**
PC/WE	0.309
PC/NW	1.27E-05
PC/WO	2.58E-05
WE/NW	1.21E-04
WE/WO	2.50E-04
NW/WO	0.649

*p-values obtained from pairwise student's t-Test
Highlighted: significant p-values ($p<0.05$)

A-3. Average Total Fatty Acid Content for Ten Cultivars

		Average Total FA Content (% DW)	Std. Dev.
Dennison	PC	18.2	1.4
	WE	17.5	0.6
	NW	17.0	0.3
	WO	17.3	1.2
H09-4	PC	15.9	1.4
	WE	16.1	1.1
	NW	15.3	0.8
	WO	14.4	1.3
M09-W043	PC	17.5	0.1
	WE	15.6	0.4
	NW	17.3	1.6
	WO	16.8	0.7
M09-W053	PC	16.1	0.4
	WE	18.9	0.5
	NW	16.1	0.5
	WO	16.3	0.3
M09-W150	PC	17.6	1.3
	WE	18.0	1.3
	NW	17.2	0.8
	WO	16.5	1.1
M09-W153	PC	16.9	1.4
	WE	17.0	1.1
	NW	15.6	1.0
	WO	16.8	1.6
Ohio FG1	PC	15.4	0.6
	WE	16.8	0.5
	NW	17.4	0.5
	WO	17.7	0.5
Ohio FG5	PC	15.3	1.0
	WE	15.4	0.9
	NW	15.7	0.3
	WO	15.9	0.7
Summit	PC	15.8	1.4
	WE	17.2	0.2
	NW	17.4	0.9
	WO	17.3	0.4
Wyandot	PC	17.5	0.2
	WE	15.2	0.9
	NW	17.5	0.8
	WO	18.3	0.3

A-4. Average Total Protein Content for Ten Cultivars

		Average Total Protein Content (%DW)	Std. Dev.
Dennison	PC	39.9	4.3
	WE	40.6	2.3
	NW	37.8	0.8
	WO	38.3	1.6
H09-4	PC	43.1	1.0
	WE	44.4	3.3
	NW	41.3	3.7
	WO	41.5	1.6
M09-W043	PC	37.2	3.1
	WE	39.3	2.7
	NW	37.2	2.3
	WO	39.0	2.3
M09-W053	PC	42.3	3.7
	WE	41.6	1.3
	NW	38.6	2.3
	WO	41.9	0.7
M09-W150	PC	36.4	2.3
	WE	37.5	3.0
	NW	36.1	1.9
	WO	35.8	1.3
M09-W153	PC	44.5	2.4
	WE	39.4	3.1
	NW	34.4	1.4
	WO	32.3	3.0
Ohio FG1	PC	45.3	3.5
	WE	43.4	1.5
	NW	35.3	2.8
	WO	34.4	2.8
Ohio FG5	PC	36.4	2.1
	WE	44.2	2.6
	NW	43.9	3.3
	WO	43.9	1.3
Summit	PC	38.5	4.0
	WE	38.4	3.8
	NW	37.3	2.2
	WO	37.5	0.7
Wyandot	PC	40.8	3.0
	WE	43.4	2.1
	NW	39.3	1.5
	WO	33.3	2.2

A-5. Oil Composition for Four Cultivars

			Content of Each Fatty Acid (% Total Oil)									
	Total FA (% DW)		C16:0	Std. Dev.	C18:0	Std. Dev.	C18:1	Std. Dev.	C18:2	Std. Dev.	C18:3	Std. Dev.
Dennison	PC	18.2	12.6	0.1	5.8	0.7	25	1.0	49.3	0.1	7.3	0.3
	WE	17.5	12.3	0.4	5.1	0.6	24.2	2.5	50.3	2.4	8.2	0.9
	NW	17	12.7	0.3	5.2	0.5	22.2	0.1	50.7	1.6	9.0	0.6
	WO	17.3	12.6	0.4	5.2	0.1	25.9	3.1	48.3	2.9	7.9	0.6
M09-W053	PC	17.2	12.6	0.8	5.3	0.4	27.2	2.6	47.5	2.2	7.4	0.6
	WE	18.9	12.5	0.2	5.1	0.3	28	0.6	47	0.6	7.4	0.3
	NW	16.1	14.9	1.0	6.0	0.9	27.9	2.1	44.2	1.4	7.0	1.0
	WO	16.3	12.1	0.4	5.0	0.0	31.5	5.1	44.0	4.0	7.3	0.8
M09-W150	PC	17.6	11.0	1.2	5	0.4	27.5	3.5	49.3	2.6	7.2	0.4
	WE	18	10.5	0.9	5.3	0.9	33.1	4.8	44.9	3.8	6.2	0.6
	NW	17.2	8.7	1.7	5.8	0.6	36.8	1.9	42.5	1.2	6.2	0.3
	WO	16.5	7.8	0.2	5.6	0.4	35.5	2.9	44.7	2.4	6.5	0.9
Wyandot	PC	17.5	11.5	0.6	6.1	0.8	25.9	1.9	49.1	1.7	7.5	0.6
	WE	15.2	11.0	0.1	7.6	0.1	25.3	0.3	48.3	0.6	7.8	0.3
	NW	17.5	11.0	0.3	5.3	0.3	27.2	3.0	48.8	2.3	7.6	0.6
	WO	18.3	11.1	0.1	5.2	0.4	28.8	3.8	48.1	3.5	6.8	0.4

A-6. p-values for Total Oil and Protein Differences for Four Cultivars

p-values		
Wyandot	**Total Oil**	**Total Protein**
PC/WE	0.011	0.213
PC/NW	0.983	0.409
PC/WO	0.008	0.008
WE/NW	0.009	0.021
WE/WO	0.004	0.001
NW/WO	0.132	0.005
Dennison	**Total Oil**	**Total Protein**
PC/WE	0.409	0.767
PC/NW	0.167	0.404
PC/WO	0.348	0.533
WE/NW	0.196	0.087
WE/WO	0.746	0.157
NW/WO	0.641	0.585
M09-W053	**Total Oil**	**Total Protein**
PC/WE	0.0001	0.722
PC/NW	0.829	0.146
PC/WO	0.343	0.821
WE/NW	0.0002	0.073
WE/WO	0.0003	0.713
NW/WO	0.521	0.057
M09-W150	**Total Oil**	**Total Protein**
PC/WE	0.654	0.577
PC/NW	0.634	0.874
PC/WO	0.241	0.649
WE/NW	0.332	0.475
WE/WO	0.123	0.337
NW/WO	0.326	0.749

*p-values obtained from pairwise student's t-Test
Highlighted: significant p-values ($p<0.05$)

A-7. Relative Average Abundances of Amino Acids for Four Cultivars

Alanine			
		Average Abundance (% Mass Abundance)	Standard Deviation
Dennison	PC	5.52	0.29
	WE	5.84	0.12
	NW	5.33	0.04
	WO	5.33	0.14
M09-W053	PC	5.18	0.39
	WE	5.24	0.31
	NW	5.26	0.09
	WO	5.27	0.07
M09-W150	PC	5.07	0.25
	WE	4.89	0.10
	NW	5.12	0.18
	WO	5.21	0.13
Wyandot	PC	5.55	0.19
	WE	5.45	0.10
	NW	5.31	0.12
	WO	5.18	0.19

Arginine			
		Average Abundance (% Mass Abundance)	Standard Deviation
Dennison	PC	5.31	0.31
	WE	5.07	0.40
	NW	5.47	0.59
	WO	5.13	0.28
M09-W053	PC	5.47	0.52
	WE	4.88	0.57
	NW	5.48	0.28
	WO	5.23	0.08
M09-W150	PC	4.77	0.24
	WE	5.08	0.29
	NW	5.14	0.57
	WO	5.25	0.21
Wyandot	PC	5.10	0.17
	WE	5.60	0.63
	NW	5.18	0.15
	WO	5.37	0.25

Aspartate			
		Average Abundance (% Mass Abundance)	Standard Deviation
Dennison	PC	8.74	0.48
	WE	8.08	0.96
	NW	8.67	0.20
	WO	8.71	0.17
M09-W053	PC	8.17	0.11
	WE	8.65	0.61
	NW	8.35	0.42
	WO	8.82	0.06
M09-W150	PC	8.87	0.47
	WE	8.72	0.23
	NW	8.39	0.13
	WO	8.48	0.11
Wyandot	PC	8.11	0.88
	WE	8.54	0.29
	NW	8.20	0.09
	WO	8.23	0.28

Glutamate			
		Average Abundance (% Mass Abundance)	Standard Deviation
Dennison	PC	13.67	0.63
	WE	12.56	0.59
	NW	11.12	0.48
	WO	11.45	0.35
M09-W053	PC	14.61	0.72
	WE	14.26	0.27
	NW	11.43	0.25
	WO	11.74	0.29
M09-W150	PC	14.59	0.47
	WE	14.67	0.45
	NW	12.85	0.08
	WO	12.17	0.40
Wyandot	PC	13.26	0.79
	WE	13.67	0.31
	NW	12.59	0.29
	WO	12.70	0.42

Glycine			
		Average Abundance (% Mass Abundance)	Standard Deviation
Dennison	PC	5.29	0.20
	WE	5.60	0.21
	NW	5.08	0.11
	WO	5.16	0.14
M09-W053	PC	5.01	0.25
	WE	4.96	0.25
	NW	4.89	0.12
	WO	5.09	0.11
M09-W150	PC	5.05	0.28
	WE	5.08	0.13
	NW	5.14	0.13
	WO	5.06	0.06
Wyandot	PC	5.46	0.16
	WE	5.17	0.17
	NW	5.15	0.15
	WO	5.13	0.11

Histidine			
		Average Abundance (% Mass Abundance)	Standard Deviation
Dennison	PC	3.26	0.12
	WE	3.27	0.12
	NW	2.95	0.09
	WO	2.95	0.05
M09-W053	PC	3.28	0.12
	WE	3.21	0.06
	NW	2.94	0.04
	WO	2.90	0.11
M09-W150	PC	3.33	0.10
	WE	3.37	0.16
	NW	2.67	0.09
	WO	2.72	0.07
Wyandot	PC	3.48	0.12
	WE	3.39	0.14
	NW	2.76	0.09
	WO	2.70	0.09

Isoleucine			
		Average Abundance (% Mass Abundance)	Standard Deviation
Dennison	PC	6.37	0.44
	WE	6.81	0.25
	NW	6.22	0.12
	WO	6.29	0.25
M09-W053	PC	5.81	0.13
	WE	5.69	0.07
	NW	6.38	0.11
	WO	6.27	0.16
M09-W150	PC	5.69	0.30
	WE	5.38	0.25
	NW	5.69	0.07
	WO	5.82	0.12
Wyandot	PC	6.74	0.41
	WE	6.48	0.35
	NW	5.86	0.06
	WO	5.66	0.26

Leucine			
		Average Abundance (% Mass Abundance)	Standard Deviation
Dennison	PC	12.21	0.50
	WE	12.52	0.18
	NW	12.96	0.34
	WO	12.81	0.12
M09-W053	PC	11.76	0.22
	WE	12.42	0.89
	NW	13.06	0.25
	WO	13.02	0.29
M09-W150	PC	12.51	0.19
	WE	11.74	0.69
	NW	11.87	0.22
	WO	11.79	0.35
Wyandot	PC	12.12	0.81
	WE	11.92	0.16
	NW	11.73	0.40
	WO	11.63	0.19

Lysine			
		Average Abundance (% Mass Abundance)	Standard Deviation
Dennison	PC	7.01	0.49
	WE	7.02	0.48
	NW	7.48	0.11
	WO	7.43	0.20
M09-W053	PC	7.42	0.66
	WE	8.28	0.88
	NW	8.06	0.14
	WO	7.87	0.14
M09-W150	PC	6.96	0.09
	WE	7.10	0.17
	NW	6.62	0.28
	WO	6.77	0.16
Wyandot	PC	6.98	0.27
	WE	6.72	0.39
	NW	6.62	0.26
	WO	6.80	0.28

Methionine			
		Average Abundance (% Mass Abundance)	Standard Deviation
Dennison	PC	1.20	0.23
	WE	1.48	0.22
	NW	1.97	0.13
	WO	2.05	0.12
M09-W053	PC	1.53	0.30
	WE	1.57	0.06
	NW	1.41	0.31
	WO	1.45	0.30
M09-W150	PC	1.17	0.26
	WE	1.44	0.09
	NW	1.52	0.11
	WO	1.48	0.35
Wyandot	PC	1.71	0.14
	WE	1.61	0.07
	NW	1.67	0.10
	WO	1.61	0.17

Phenylalanine			
		Average Abundance (% Mass Abundance)	Standard Deviation
Dennison	PC	5.82	0.12
	WE	6.05	0.21
	NW	5.09	0.13
	WO	5.14	0.06
M09-W053	PC	6.30	0.41
	WE	5.89	0.45
	NW	5.27	0.14
	WO	5.30	0.17
M09-W150	PC	6.28	0.46
	WE	6.34	0.39
	NW	7.95	0.39
	WO	8.21	0.36
Wyandot	PC	6.18	0.26
	WE	6.21	0.19
	NW	8.38	0.25
	WO	8.16	0.17

Proline			
		Average Abundance (% Mass Abundance)	Standard Deviation
Dennison	PC	5.27	0.19
	WE	5.22	0.37
	NW	5.64	0.14
	WO	5.62	0.23
M09-W053	PC	5.11	0.22
	WE	4.87	0.26
	NW	5.59	0.19
	WO	5.67	0.07
M09-W150	PC	4.73	0.16
	WE	4.70	0.18
	NW	4.90	0.12
	WO	4.81	0.19
Wyandot	PC	5.10	0.38
	WE	5.06	0.19
	NW	4.60	0.09
	WO	4.48	0.09

Serine			
		Average Abundance (% Mass Abundance)	Standard Deviation
Dennison	PC	6.07	0.36
	WE	6.20	0.24
	NW	6.93	0.12
	WO	6.84	0.15
M09-W053	PC	6.06	0.18
	WE	6.07	0.41
	NW	6.62	0.09
	WO	6.77	0.15
M09-W150	PC	6.49	0.22
	WE	6.60	0.32
	NW	6.30	0.68
	WO	6.15	0.06
Wyandot	PC	5.61	0.34
	WE	5.66	0.13
	NW	6.04	0.61
	WO	6.60	0.18

Threonine			
		Average Abundance (% Mass Abundance)	Standard Deviation
Dennison	PC	4.45	0.23
	WE	4.00	0.31
	NW	4.79	0.24
	WO	4.94	0.09
M09-W053	PC	4.29	0.39
	WE	4.36	0.13
	NW	4.69	0.18
	WO	4.67	0.04
M09-W150	PC	4.37	0.10
	WE	4.08	0.13
	NW	4.68	0.14
	WO	4.63	0.30
Wyandot	PC	4.33	0.19
	WE	4.22	0.16
	NW	4.31	0.22
	WO	4.49	0.23

Tyrosine			
		Average Abundance (% Mass Abundance)	Standard Deviation
Dennison	PC	3.59	0.23
	WE	3.79	0.16
	NW	3.62	0.08
	WO	3.62	0.06
M09-W053	PC	4.52	0.33
	WE	4.33	0.24
	NW	3.57	0.08
	WO	3.45	0.06
M09-W150	PC	4.38	0.11
	WE	4.40	0.25
	NW	5.20	0.14
	WO	5.34	0.23
Wyandot	PC	4.03	0.18
	WE	3.81	0.14
	NW	5.59	0.11
	WO	5.40	0.19

Valine			
		Average Abundance (% Mass Abundance)	Standard Deviation
Dennison	PC	6.22	0.24
	WE	6.50	0.39
	NW	6.66	0.10
	WO	6.54	0.05
M09-W053	PC	6.03	0.23
	WE	5.84	0.09
	NW	6.68	0.11
	WO	6.66	0.13
M09-W150	PC	5.74	0.35
	WE	5.88	0.31
	NW	5.95	0.29
	WO	6.12	0.15
Wyandot	PC	6.24	0.36
	WE	6.49	0.14
	NW	6.00	0.12
	WO	5.88	0.06

References

1. Alonso, A.P., et al., *Carbon conversion efficiency and central metabolic fluxes in developing sunflower (Helianthus annuus L.) embryos.* Plant Journal, 2007. **52**(2): p. 296-308.

2. Alonso, A.P., V.L. Dale, and Y. Shachar-Hill, *Understanding fatty acid synthesis in developing maize embryos using metabolic flux analysis.* Metabolic Engineering, 2010. **12**(5): p. 488-497.

3. Bachlava et al., *Heritability of oleic acid content in soybean seed oil and its genetic correlation with fatty acid and agronomics traits.* Crop Science, 2008a. **48:** p. 1764-1772.

4. Blau, N., Van Spronsen, F.J., Levy, H.L. *Phenylketonuria*. The Lancet, 2010. **376** (9750): p. 1417-1427. Doi: 10.1016/S0140-6736(10)60961-0

5. Clemente, T.E. and E.B. Cahoon, *Soybean oil: genetic approaches for modification of functionality and total content.* Plant physiology, 2009. **151**(3): p. 1030-40.

6. Cocuron, J.C. and A.P. Alonso, *Liquid chromatography tandem mass spectrometry for measuring 13C-labeling in intermediaries of the glycolysis and pentose-phosphate pathway.* Methods in Molecular Biology, 2014. **1090**: p. 131-142.

7. Cocuron, J.C., et al., *Targeted Metabolomics of Physaria fendleri, an Industrial crop Producting Hydroxy Fatty Acids.* Plant & Cell Physiology, 2014. **55(3)**: p. 620-633.

8. Denicoff, Marina R., Marvin Prater, and Pierre Bahizi, Soybean Transportation Profile. U.S. Department of Agriculture, Agricultural Marketing Service, October 2014. Web. http://dx.doi.org/10.9752/TS203.10-2014

9. Dorff, E. *The Soybean, agriculture's jack-of-all-trades, is gaining ground across Canada.* Statistics Canada. Web. http://www.statcan.gc.ca/pub/96-325-x/2007000/article/10369-eng.htm. Accessed 4/28/2014, 2014.

10. Fehr WR, Welke GA, Hammond EG, Duvick DN, Cianzio SR (1992) *Inheritance of reduced linolenic acid content in soybean genotypes A16 and A17*. Crop Sci, 1992. **32** p. 903–906.

11. Goffman, F.D., et al., *Light enables a very high efficiency of carbon storage in developing embryos of rapeseed.* Plant Physiology, 2005. **138**(4): p. 2269-2279.

12. Han Q, Feng H, Zhao H, Huang L, Wang X, et al. (2013) *Effect of a benzothiadiazole on inducing resistance of soybean to Phytophthora sojae. Protoplasma* 250: 471–481. doi: 10.1007/s00709-012-0430-6

13. James, C. *Global Status of Commercialized Biotech/GM Crops*, 2014. ISAAA Brief No. 49. ISAAA: Ithaco, NY.

14. Joshi T., Valliyodan B., Wu J., Lee S., Xu D. and Nguyen H. *Genomic differences between cultivated soybean, G. max and its wild relative G. soja*. BMC Genomics 2013;14:S5.

15. Kazmierczak, S. *Diseases of Metabolism (Disorders of Amino Acid Metabolism)*. Analytical Chemistry, 1993. **65**(12): p. 401-404.

16. Krishnan H.B., *Engineering Soybean for Enhanced Sulfur Amino Acid Content*. Crop Sci 2005/3;45(2): p. 454-461.

17. Lee, H., Park, S. Y., Zhang, Z.J. *An Overview of Genetic Transformation of Soybean*. En: Board, J.W.A Comprehensive Survey of International Soybean Research – Genetics, Physiology, Agonomy and Nitogen Relationships. Ed. Intech. 2013. p. 489-506. ISBN: 978-953-51-0876-4.

18. Marega Filho, M.; Destro, D.; Miranda, L. A.; Spinosa, W. A; Carrão-Panizzi, M. C.. and Montalván, R. (2001), Relationships among oil content, protein content and seed size in soybeans. *Braz. arch. biol. technol.*, **44**, 23-32.

19. McHale, L. *Breeding Ohio-Adapted Soybean Cultivars*. Available at: http://hcs.osu.edu/mchalelab/projects/breeding-ohio-adapted-soybean-cultivars. Accessed March, 2014.

20. Medic J, A.C., Hurburgh CH, *Current Knowledge in Soybean Composition.* Journal of the American Oil Chemists' Society, 2013: p. DOI:10.1007/s11746-013-2407-9

21. National Soybean Research Laboratory. *Production Basics.* Available at: http://nsrl.illinois.edu/content/production-basics. Accessed March, 2014.

22. Penning, B.W., et al., *Genetic resources for maize cell wall biology.* Plant physiology, 2009. **151**(4): p. 1703-28.

23. Qiu L. and Chang R. The Origin and History of Soybean. In: Singh G, editor. Soybean : Botany, Production and Uses Wallingford, Oxon, GBR: CABI Publishing; 2010. p. 1-2.

24. Rahman et al., *A novel soybean germplasm with elevated saturated fatty acids*. Crop Science, 2003. **43:** p. 527-531.

25. Rahman et al., *Combining ability in loci for high oleic and low linolenic acids in soybean.* Crop Science, 2001. **41**: p. 26-29.

26. Rotundo, J.L. and Westgate, M.E. *Meta-analysis of environmental effect on soybean seed composition.* Field Crops Research, 2009. **110**(2): p. 147-156.

27. Shurtleff, W., and Aoyagi, A. Early Named Soybean Varieties in the United States and Canada. Lafayette, CA: Soyinfo Center, 2013. Soyinfo Center, 12 Oct. 2013. Web. 3 Apr. 2015.

28. Sleper, D.A. & Shannon, J.G. (2003). Role of public and private soybean breeding programs in the development of soybean varieties using biotechnology. *AgBioForum*, *6*(1&2), 27-32. Available on the World Wide Web: http://www.agbioforum.org.

29. Streeter J.G., *Effects of drought on nitrogen fixation in soybean root nodules*. Plant, Cell Environ 2003;26(8): p. 1199-1204.

30. Takagi, Y. and Rahman S.M., *Inheritance of high oleic acid content in the seed oil of soybean mutant M23*. Theoretical and Applied Genetics, 1996. **92**: p.179-182.

31. Texas A&M University Protein Chemistry Lab. *Amino Acid Analysis – Quantitation and Data Analysis.* Available at: http://tamupcl.com/Services/AminoAcidAnalysis/AAAQuantitationandDataAnalysis. Accessed April, 2015.

32. Trumbo P, Schlicker S, Yates AA, Poos M; Food and Nutrition Board of the Institute of Medicine, The National Academies. Dietary reference intakes for energy, carbohydrate, fiber, fat, fatty acids, cholesterol, protein and amino acids. J Am Diet Assoc. 2002;102(11):1621-1630.

33. Tsogtbaatar, E. et al., *Metabolite fingerprinting of pennycress (Thlaspi arvense L.) embryos to assess active pathways during oil synthesis.* Journal of Experimental Botany, 2015: Doi: 10.1093/jxb/erv020

34. USDA ERS - Soybeans & Oil Crops. Available at: http://www.ers.usda.gov/topics/crops/soybeans-oil-crops.aspx. Accessed February-March, 2015.

35. USDA ERS – Oil Crops Yearbook. Available at: http://www.ers.usda.gov/data-products/oil-crops-yearbook.aspx. Accessed March-April, 2015.

36. Wang T.L., Domoney C., Hedley C.L., Casey R., Grusak M.A. *Can We Improve the Nutritional Quality of Legume Seeds?* Plant Physiology 2003 March 01;131(3): p. 886-891.

37. Xia, J., Mandal, R., Sinelnikov, I., Broadhurst, D., and Wishart, D.S. (2012) MetaboAnalyst 2.0 - a comprehensive server for metabolomic data analysis . Nucl. Acids Res. 40, W127-W133.

38. Xia, J., Psychogios, N., Young, N. and Wishart, D.S. (2009) MetaboAnalyst: a web server for metabolomic data analysis and interpretation. Nucl. Acids Res. 37, W652-660.